Общи задачи

Тодор Дуков

Примери за задачи с числа

Искаме да напишем алгоритъм, който при подадена двойка от естествени числа $\langle a, b\rangle$, където $b \leq a$, да върне $\gcd(a, b)$ т.е. това число $d \in \mathbb{N}$, за което:

- $d \mid a$ и $d \mid b$
- ако $d_1 \mid a$ и $d_1 \mid b$, то $d_1 \mid d$

За целта ще покажем, че за всяко $a, b, q, r \in \mathbb{N}$, където $0 < b \leq a$, $a = bq + r$ и $r \in \{0, \ldots, b-1\}$, е изпълнено, че:

$$\gcd(a, b) = \gcd(b, r)$$

Първо имаме, че $\gcd(a, b) \mid a$ и $\gcd(a, b) \mid b$, откъдето понеже $a = bq+r$, имаме $\gcd(a, b) \mid r$. Тогава $\gcd(a, b) \mid \gcd(b, r)$. От друга страна $\gcd(b, r) \mid b$ и $\gcd(b, r) \mid r$, откъдето $\gcd(b, r) \mid bq + r = a$. Така $\gcd(b, r) \mid \gcd(a, b)$. Накрая можем да заключим $\gcd(a, b) = \gcd(b, r)$ от $\gcd(a, b) \mid \gcd(b, r)$ и $\gcd(b, r) \mid \gcd(a, b)$.
На базата на това наблюдение се получава следният алгоритъм (кръстен на Евклид):

```
int euclid(int a, int b) // a >= b
{
    if (b == 0)
        return a;

    return euclid(b, a % b);
}
```

Ще покажем с индукция относно $\mathtt{b}$, че за всяко $\mathtt{a} \geq \mathtt{b}$, е изпълнено, че:

$$\mathtt{euclid}(\mathtt{a}, \mathtt{b}) = \mathtt{gcd}(\mathtt{a}, \mathtt{b})$$

- В базовия случай имаме $\mathtt{euclid}(\mathtt{a}, 0) = \mathtt{a} = \mathtt{gcd}(\mathtt{a}, 0)$
- Нека $\mathtt{b} > 0$ и $\mathtt{a} = \mathtt{bq} + \mathtt{r}$ за някои $\mathtt{r} \in \{0, \ldots, \mathtt{b} - 1\}$ и $\mathtt{q}$ и нека твърдението е изпълнено за всяко $\mathtt{b}' < \mathtt{b}$. Тогава:

$$\mathtt{euclid}(\mathtt{a}, \mathtt{b}) = \mathtt{euclid}(\mathtt{b}, \mathtt{r}) \overset{\text{(ИП)}}{=} \mathtt{gcd}(\mathtt{b}, \mathtt{r}) = \mathtt{gcd}(\mathtt{a}, \mathtt{b})$$

Алгоритъмът терминира, понеже управляващият параметър $\mathtt{b}$ винаги намаля, докато не стане $\mathtt{0}$. На пръсти ще покажем, че алгоритъмът има сложност по време и памет $O(\log(a))$. Нека видим, че ако $a > b$, то $\mathrm{mod}(a, b) < \frac{a}{2}$:

1 сл. ако $b \leq \frac{a}{2}$, то $\mathrm{mod}(a, b) < \frac{a}{2}$ и сме готови

2 сл. ако пък $b > \frac{a}{2}$, то тогава $a - b < \frac{a}{2}$, откъдето $\mathrm{mod}(a, b) = \mathrm{mod}(a - b, b) = a - b < \frac{a}{2}$

Тогава през на всеки две стъпки на алгоритъма от вход $\langle a, b\rangle$, отиваме до вход $\langle \mathrm{mod}(a, b), \mathrm{mod}(b, \mathrm{mod}(a, b))\rangle$ и левият аргумент става поне два пъти по-малък. Тъй като левият аргумент е горна граница за десния, то той също ще намаля рязко. Дълбочината на дървото на рекурсията зависи само от броя на рекурсивните извиквания. Понеже те са логаритмично много и всяко едно от тях заема константна памет, сложността по памет е също $O(\log(a))$.
Втората задача за числа е скрита в задача за масиви:
Имаме един масив $\mathtt{A}[\mathtt{1} \ldots \mathtt{n}]$, който съдържа всички числа от $\mathtt{1}$ до $\mathtt{n}$, с изключение на едно от тях, което е заместено с друго число от $\mathtt{1}$ до $\mathtt{n}$. Искаме да напишем колкото се може по-бърз алгоритъм, който при вход такъв масив $\mathtt{A}$ връща наредената двойка $\langle \mathtt{dup}, \mathtt{miss}\rangle$ от дупликата и липсващото число. Оказва се, че тази задача може да се реши за линейно време и константна памет. За целта трябва да се забележат следните факти:

- $\mathtt{dup} - \mathtt{miss} = \sum\limits_{\mathtt{i}=1}^{\mathtt{n}} \mathtt{A}[\mathtt{i}] - \sum\limits_{\mathtt{i}=1}^{\mathtt{n}} \mathtt{i} = \sum\limits_{\mathtt{i}=1}^{\mathtt{n}} \mathtt{A}[\mathtt{i}] - \frac{\mathtt{n}(\mathtt{n}+1)}{2} =: \mathtt{S}_{1,\mathtt{A}}$
- $(\mathtt{dup} + \mathtt{miss})(\mathtt{dup} - \mathtt{miss}) = \mathtt{dup}^2 - \mathtt{miss}^2 = \sum\limits_{\mathtt{i}=1}^{\mathtt{n}} \mathtt{A}[\mathtt{i}]^2 - \sum\limits_{\mathtt{i}=1}^{\mathtt{n}} \mathtt{i}^2 = \sum\limits_{\mathtt{i}=1}^{\mathtt{n}} \mathtt{A}[\mathtt{i}]^2 - \frac{\mathtt{n}(\mathtt{n}+1)(2\mathtt{n}+1)}{6} =: \mathtt{S}_{2,\mathtt{A}}$

От тях получаваме следната система от уравнения:

$$\left|\begin{array}{l} \texttt{dup} - \texttt{miss} = \texttt{S}_{\texttt{1,A}} \\ \texttt{dup} + \texttt{miss} = \frac{\texttt{S}_{\texttt{2,A}}}{\texttt{S}_{\texttt{1,A}}} \end{array}\right.$$

Най-сложното, което трябва да направим, е да пресметнем $\texttt{S}_{\texttt{1,A}}$ и $\texttt{S}_{\texttt{2,A}}$:

```
1  pair<int, int> find_missing_and_duplicate(int *arr, int n)
2  {
3      int sum = (n * (n + 1)) / 2;
4      int sq_sum = (n * (n + 1) * (2 * n + 1)) / 6;
5
6      int arr_sum = 0;
7      int arr_sq_sum = 0;
8
9      for (int i = 0; i < n; ++i)
10     {
11         arr_sum += arr[i];
12         arr_sq_sum += arr[i] * arr[i];
13     }
14
15     int dup = ((arr_sum - sum) + (arr_sq_sum - sq_sum) / (arr_sum - sum)) / 2; // (S1 + S2 / S1) / 2
16     int miss = dup - arr_sum + sum; // dup - S1
17
18     return {dup, miss};
19 }
```

Остава само да се докаже следното твърдение:

Инвариант. *При всяко достигане на проверката за край на цикъла на ред 9 имаме, че:*

$$\texttt{arr\_sum} = \sum_{\texttt{k=0}}^{\texttt{i}-1} \texttt{arr[k]} \textit{ и } \texttt{arr\_sq\_sum} = \sum_{\texttt{k=0}}^{\texttt{i}-1} \texttt{arr[k]}^2$$

База. При първото достигане имаме, че:

- $\texttt{arr\_sum} = 0 = \sum_{\texttt{k=0}}^{0-1} \texttt{arr[k]}$
- $\texttt{arr\_sq\_sum} = 0 = \sum_{\texttt{k=0}}^{0-1} \texttt{arr[k]}^2$

Поддръжка. Нека твърдението е изпълнено за някое непоследно достигане на проверката за край на цикъла. Тогава влизайки в тялото на цикъла:

- $\texttt{arr\_sum}$ става $\texttt{arr\_sum} + \texttt{arr[i]} \overset{\text{(ИП)}}{=} \sum_{\texttt{k=0}}^{\texttt{i}-1} \texttt{arr[k]} + \texttt{arr[i]} = \sum_{\texttt{k=0}}^{\texttt{i}} \texttt{arr[k]} = \sum_{\texttt{k=0}}^{\texttt{i}_{\texttt{new}}-1} \texttt{arr[k]}$
- $\texttt{arr\_sq\_sum}$ става $\texttt{arr\_sq\_sum} + \texttt{arr[i]}^2 \overset{\text{(ИП)}}{=} \sum_{\texttt{k=0}}^{\texttt{i}-1} \texttt{arr[k]}^2 + \texttt{arr[i]}^2 = \sum_{\texttt{k=0}}^{\texttt{i}} \texttt{arr[k]}^2 = \sum_{\texttt{k=0}}^{\texttt{i}_{\texttt{new}}-1} \texttt{arr[k]}^2$

Терминация. В последното достигане на проверката за край на цикъла имаме, че $\texttt{i} = \texttt{n}$, откъдето:

$$\texttt{arr\_sum} = \sum_{\texttt{i=0}}^{\texttt{n}-1} \texttt{arr[i]} \text{ и } \texttt{arr\_sq\_sum} = \sum_{\texttt{i=0}}^{\texttt{n}-1} \texttt{arr[i]}^2$$

Имайки това твърдение, остава само да се отбележи, че $\texttt{dup}$ и $\texttt{miss}$ наистина са решения на горната система от уравнения (със съответните преиндексирания). Накрая нека за пълнота с индукция по $n \in \mathbb{N}$ покажем, че:

$$\sum_{i=0}^{n} i = \frac{n(n+1)}{2} \text{ и } \sum_{i=0}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$$

- $\sum_{i=0}^{0} i = 0 = \frac{0(0+1)}{2}$ и $\sum_{i=0}^{0} i^2 = \frac{0(0+1)(2\cdot 0+1)}{6}$ ✓
- $\sum_{i=0}^{n+1} i = \sum_{i=0}^{n} i + (n+1) \overset{\text{(ИП)}}{=} \frac{n(n+1)}{2} + (n+1) = \frac{n(n+1)}{2} + \frac{2(n+1)}{2} = \frac{(n+1)(n+2)}{2}$
- $\sum_{i=0}^{n+1} i^2 = \sum_{i=0}^{n} i^2 + (n+1)^2 \overset{\text{(ИП)}}{=} \frac{n(n+1)(2n+1)}{6} + (n+1)^2 = \frac{n(n+1)(2n+1)}{6} + \frac{6(n+1)^2}{6} = \frac{(n+1)(2n^2+n+6n+6)}{6} = \frac{(n+1)(n+2)(2(n+1)+1)}{6}$